Jan 19, 2025

Pre-Analysis Plan: An Evaluation of a Remedial Summer Camp in the Dominican Republic

Carolina Concha-Arriagada, Guthrie Gray-Lobe, and Michael Kremer

1. Introduction

a. Abstract

This study evaluates the cost-effectiveness of a three-week (15-day) summer camp program in the Dominican Republic to reduce grade repetition and improve academic performance for students struggling with foundational literacy and numeracy. The program, developed and implemented by the Dominican Ministry of Education (MINERD), targets low-achieving and overage students in grades 3-5 who are at high risk of being held back. We will examine the impact of invitations to the program on attendance at the camp, test scores, grade retention, and dropout using randomized invitations to the program. We have already examined the impacts on attendance. Randomized invitations are associated with 3.9 more days of summer camp attendance and a 26 percentage point increase in the fraction of students who attend at least two-thirds (10 days) of the course. We will also examine the impact of attendance at the camp directly on test scores, grade retention, and dropout.

b. Intervention

The summer camp program was a three-week (15-day) intensive intervention designed to improve foundational literacy and numeracy skills among primary school students (grades 3-5) in the Dominican Republic. The Dominican Ministry of Education (MINERD) developed and implemented the program from July 8th to July 26th, 2024.

The program targeted low-achieving and overage students at high risk of being held back in the next year. Students in the program were divided into grade-specific classrooms of 25, with two teachers assigned to each classroom. Daily activities were held from 8:30 am to 3:30 pm.

The program focused on strengthening students' foundational skills in literacy and numeracy, aligned with the national curriculum. The specific content and activities included:

- Literacy: Developing phonemic awareness, phonics skills, reading fluency, and comprehension strategies.
- Numeracy: Reinforcing number sense, basic operations (addition, subtraction, multiplication, division), and problem-solving skills.

The program utilized various instructional methods, including interactive games and activities, small group instruction to provide targeted support and address individual learning needs, and individualized practice to reinforce concepts and build fluency.

MINERD has expressed interest in implementing another summer camp program, at least equal in size, in mid-2025. Results of the study may contribute to program design refinements. We anticipate that the program's next iteration will allow for out-of-sample testing of results from the 2024 round.

c. Research Questions

This study seeks to answer the following research questions:

- How does the 3-week intensive summer camp impact students' academic performance in literacy and numeracy?
- Does the summer camp program reduce grade repetition?
- Is the program more effective for some particular subgroups of students (i.e., by Haitian and Dominican nationality, gender, distance from the student's school to the camp, or grade)? We discuss concerns regarding statistical power for sub-group analysis.

2. *Research Strategy*

a. *Sampling*

i. Sampling Frame

The eligible population for the study consists of public primary school students (grades 3-5) who have the lowest performance in their classroom from 30 educational districts in the Dominican Republic. The government initially identified the subset of schools participating in the program based on an assessment of their characteristics (number of overage students, repetition rate, students' average scores in math and language). MINERD asked teachers in study schools to nominate the students with the lowest grades to identify these students.

The study sample includes 5,482 students: 3,369 students assigned to the summer camp treatment group and 2,113 students in the comparison group.

ii. Statistical Power

In this section, we report the minimum detectable effect of the intent-to-treat effect of inviting students to come to the remedial camp. We focus on intent-to-treat effects for

now given econometric challenges with identifying the causal effect of a particular measure of enrollment given the heterogeneous effect of invitations on summer camp attendance discussed below.

Power calculations for test scores assume that baseline test scores and stratifying covariates will absorb up to 40 percent of the variation in test scores. This figure may be conservative given that the baseline and endline are mirror tests.

Base rates for grade retention after one year come from national statistics.[1] Because the program targeted areas with higher grade retention, these rates will tend to be lower than for the sample. We construct assumed base rates for an indicator measuring whether a student is ever retained over three years by multiplying the one-year rate by three.

Table 1 reports the minimum detectable effects for the intent-to-treat effect for the aggregate sample and sub-groups.

Table 1. Power calculation estimates.

	Total Students	Invited students	Base rate	MDE	Unit
Aggregate					
Child outcomes (standardized test scores)	5,482	3,369	1.00	0.05	sd
Retention in grade (one year)	5,482	3,369	0.07	0.02	pp
Retention in grade (three years)	5,482	3,369	0.20	0.02	pp
Haitian students only					
Standardized test scores	583	358	1.00	0.17	sd
Retention in grade (one year)	583	358	0.07	0.06	pp

[1] Calculated using national grade retention rate 2016-2017. Source: MINERD.

	Retention in grade (three years)	583	358	0.20	0.07	pp
Female students only						
	Standardized test scores	1,813	1,115	1.00	0.09	sd
	Retention in grade (one year)	1,813	1,115	0.04	0.03	pp
	Retention in grade (three years)	1,813	1,115	0.12	0.03	pp
Male students only						
	Standardized test scores	3,669	2,254	1.00	0.07	sd
	Retention in grade (one year)	3,669	2,254	0.09	0.03	pp
	Retention in grade (three years)	3,669	2,254	0.27	0.03	pp
Grade 3						
	Standardized test scores	2,420	1,479	1.00	0.08	sd
	Retention in grade (one year)	2,420	1,479	0.08	0.03	pp
	Retention in grade (three years)	2,420	1,479	0.25	0.04	pp
Grade 4						
	Standardized test scores	1,636	1,009	1.00	0.10	sd
	Retention in grade (one year)	1,636	1,009	0.06	0.03	pp
	Retention in grade (three years)	1,636	1,009	0.19	0.04	pp
Grade 5						
	Standardized test scores	1,426	881	1.00	0.11	sd
	Retention in grade (one year)	1,426	881	0.05	0.03	pp
	Retention in grade (three years)	1,426	881	0.16	0.04	pp

The study appears to be adequately powered to test the central hypothesis of the program: that a short dose of remedial education can produce much of the benefit found in other studies of remedial and adaptive learning for struggling students. Depending on assumptions regarding the dose-response relationship between remedial education and test score outcomes, the study may be adequately powered to detect impacts comparable to those observed in the literature on Teaching at the Right Level (TaRL) (see Banerjee et al. 2017 for a review). TaRL interventions typically include more instructional time than the remedial summer camp studied here. Some of the most impactful TaRL interventions have been 50 days long, with each day consisting of 3 hours of instruction for 150 hours. Treatment-on-treated (TOT) effects for these interventions have been around 0.75 standard deviations.[2] Assuming linearity in the dose-response relationship, 60 hours of summer camp exposure would be associated with approximately 0.28 standard deviation effect. For comparison, we consider power for a test of the effect of attending 60 hours of the remedial summer camp intervention, both for a binary endogenous program compliance definition (attending at least 10 days of the intervention) and for a continuous measure of hours of attendance. Random invitations to the program increased attendance of at least 10 days by 0.258 and increased the number of total hours (attended by 3.87 days x 6 hours/day / (60 hours for “full exposure”)= 0.39. 2SLS estimates using these first stages will have an MDE of 0.22 and 0.15, which are both smaller than the 0.28 standard deviation effect from 60 hours of TaRL exposure.

Based on existing aggregate statistics, we view the study as being underpowered to detect what would be moderate reductions in grade repetition after one year. The MDE

[2] Note that the largest impacts from individual studies may tend to overstate the benefit of an intervention (see Angrist & Meager 2023).

for the effect using the full sample exceeds a 25% reduction in 1-year grade retention. For some sub-groups, especially students, the MDE is nearly equivalent to the base rate, including Haitian students, female students, and grade 5 students.

Base rates for grade retention for the sample of students in the study may be substantially higher for the study sample, given that eligible students were identified by teachers as being at higher risk of being retained in grade. We will revisit power calculations after observing realized base rates for the sample of study recipients; however, based on the available data and due to the low first stage effect of invitations on attendance, we view the study as unable to rule out what would be meaningful improvements in grade retention.

The impacts on three-year grade retention are more promising, although we note that for girls and older-grade students, the MDE's still represent more than a 25% reduction in the grade retention rate.

We discuss extensions of the analysis that will improve on the precision of our main estimates at the expense of additional assumptions in section 3.c.ii.

iii. Assignment to Treatment

The source of exogenous variation in the study is the random assignment of students to either the summer camp or the comparison group. Students were assigned to

treatment and control conditions through a stratified randomization process conducted by MINERD. The stratification variables were school and grade level.

b. Fieldwork

i. Instruments

The study will utilize the following data collection instruments:

- MINERD-developed foundational learning assessments (baseline and endline) to measure students' academic performance in literacy and numeracy. MINERD developed these assessments based on national curriculum standards.
- A demographic survey (at baseline) was used to collect information on students' background characteristics. It collects information on household composition, including the presence of parents, siblings, and other relatives, access to resources such as electricity, clean water, and sanitation facilities, and one question about who helps students with their homework.

ii. Data Collection

The entire data collection took place between June and November 2024.

Data collection entails the following activities:

- Baseline assessment and a demographic survey administered by teachers in participating schools (June 2024).
- Summer camp program implementation (July 2024).
- Endline assessments by teachers (September 2024).
- Collection of administrative data from MINERD (ongoing).
- Digitization of students' assessments and surveys by MINERD (November 2024).

We may follow up in subsequent years and update this plan as necessary.

iii. Data Processing

Data processing will involve:

- Data entry and digitization, since students filled out paper questionnaires.
- Data cleaning and validation.
- Data anonymization and de-identification.
- Data storage in encrypted folders at the University of Chicago.

Data confidentiality will be maintained through encryption, password protection, and secure data transfer protocols. After the study, the data will be archived securely and potentially used for future research according to IRB approvals.

3. Empirical Analysis

a. Variables

Outcome Variables:

- Endline assessment results in literacy and numeracy (continuous).
- Repetition (binary): Whether the student repeated a grade since the start of the intervention.
- Dropout (binary): Whether the student dropped out of school.
- Grades in sequential years (continuous): This could be collected in follow-up studies.

Treatment Variables:

- Exogenous:

    - Random assignment of students (binary): Whether the student was randomly assigned to either the summer camp or the comparison group
- Endogenous:
    - Summer camp attendance: Number of days (0-15) the student attended the summer camp program. We also consider attendance as a study outcome we have already observed.

Control Variables:

- Grade level
- Interactions between baseline test scores and grade level

For students with missing baseline test scores, we will impute the score to be the mean. Controls will include an indicator for baseline test score missingness that will be fully interacted with grade-level indicators.

b. Balancing Checks

To check for balance between the treatment and control groups, we will conduct t-tests for continuous variables and chi-squared tests for categorical variables. The variables included in these balancing checks will be baseline test score, grade level, gender, nationality, household composition, and access to resources.

To check for balance between attritors and non-attritors, we will use the same tests as above, comparing students who completed the endline assessment with those who did not. Finally, we will check for balance between attendees and non-attendees, depending on the defined threshold.

c. *Treatment Effects*

i. Intent to Treat

We will use linear regression models for continuous and binary outcomes to estimate the effect of being offered the program. The specification will include the treatment variable (summer camp offer) and the control variables listed in section 3.a.

ii. Treatment on the Treated

Analysis of the impact of participation in the remedial camps will be complicated because the program had heterogeneous effects on attendance. First, we note that there was two-sided non-compliance. Some uninvited students attended the camps and some invited students declined to attend.[3] Furthermore, attendance varies considerably, even among those who attended some of the camp. We will report impacts from a 2SLS model that assumes the dose-response relationship between attendance and test scores is approximately linear. Invited students attended six days of the camp on average, 3.87 days more than the comparison group. Forty percent of invited participants attended 10 or more days, 25.8 percentage points more than the comparison group.

To further improve precision we will examine how the ITT effect varies with the first stage observed at the remedial summer camp center level. Preliminary analysis suggests there is considerable variation in the effect of invitations on the number of

[3] Several factors contributed to uninvited students attending the summer camp. Miscommunication and confusion surrounding the selection process played a significant role. While parents were informed that not all students would be selected, the notification of who was selected was often delayed. Additionally, some camps received incorrect or incomplete lists of invited students, differing from the randomized lists. This lack of clarity and a lack of specific guidelines on handling uninvited students led some parents to send their children to the camp regardless of their official selection status. Furthermore, transportation challenges may have disproportionately affected attendance among students whose schools were located further from the camp sites, potentially leading to excluding invited students and including students not initially invited.

days attended across centers. We will first compare the first stage effect on days attended and the ITT effect on academic outcomes in a visual-instrumental-variables (VIV) scatterplot. This analysis will be complemented by analogous estimation of the overidentified instrumental variable model of the effect of days of attendance using instruments formed by the interaction of invitation status with the center to which the student was invited.

We will examine the OLS relationship between test scores and attendance at the summer camp, controlling for baseline test scores. Estimates of the effect of attendance using OLS can be biased due to selection effects, even after controlling for baseline performance. However, studies of value-added have found that controls for lagged test scores can produce unbiased estimates of causal effects of educational inputs (such as schools and teachers). Following this literature we will also compare estimates of camp attendance derived from instrumental variables and those obtained by OLS, viewing the IV estimates as noisy, but unbiased, and the OLS estimates as precise but potentially biased due to selection (Angrist et al. 2020).

The linear dose-response assumption is strong and these estimates will be considered approximations. We will try to assess the linear dose-response assumption by examining (a) the cross-sectional relationship between test scores and the number of days that a student attended the program and (b) the variation in the ITT effect and the first stage effect on days attended between groups of people who were more or less likely to attend the program. We will examine first stage and ITT variation across groups defined by the camp location where the student was assigned, the distance between their school and the camp's location. The evidence provided by these analyses will be at best speculative. Cross-sectional relationships can be biased by selection and measurement error and correlation between the ITT and first-stage effect can reflect

selection on gains when the effect of additional exposure to the program is heterogeneous.

d. Heterogeneous Effects

We will examine whether the ITT and 2SLS effects of the program vary with the following student characteristics:

- Baseline level: we will test whether the program's effect varies with their initial level.
- Gender: Prior research suggests potential differences in learning and academic performance between boys and girls. Attendance also might vary by gender.
- Age-for-grade: we will test whether the program's effect varies depending on when a student is more overage for their grade.
- Nationality: Immigrant students may face unique challenges that could influence their response to the program. Although we have already seen similar program take-up among migrants compared to Dominicans, their differential experience during the camp could impact their learning. For example, we consider that the program's content was in Spanish and that migrants may have higher transportation costs.
- Grade level: Preliminary evidence suggests that the learning levels of participating students varied considerably. The program's material, which was the same for all grade levels, may not have been optimally adapted to the students' heterogeneous needs. Notably, lower attendance rates were observed among older students, potentially indicating a lack of challenge or relevance for this group. Some 5th-grade students expressed that certain activities were not engaging, potentially due to real or perceived lack of alignment with their developmental stage. This may have contributed to lower levels of participation

among older students. Additionally, and perhaps surprisingly, teachers observed that literacy challenges were more prevalent in 4th and especially 5th grade, which they attributed to the fact that 3rd graders had been benefiting from the MINERD “Con Base” program that specifically targets literacy development. Consequently, the evaluation will incorporate a detailed analysis of treatment effects, examining how the program's impact varies across different student learning levels. We have already seen a lower attendance in grades 4th and 5th compared to 3rd.

- Student distance from camp (measured as the distance from the student's school to the remedial camp location).

We will explore these heterogeneous effects by including an interaction between the treatment variable and each of the abovementioned characteristics in the regression.